Evolution and musical rhythm

Ian Cross, Centre for Music & Science, Faculty of Music,
University of Cambridge, Cambridge, UK. Email:
ic108@cam.ac.uk (Corresponding author)

John Bispham, Centre for Music & Science, Faculty of
Music/Leverhulme Centre for Evolutionary Studies, University
of Cambridge, Cambridge, UK.

Tommi Himberg, Centre for Music & Science, Faculty of
Music, University of Cambridge, Cambridge, UK.

Joel Swaine, Centre for Music & Science, Faculty of Music,
University of Cambridge, Cambridge, UK.

Abstract: This paper explores the evolution of the human ability to perceive and behave rhythmically, particularly in interactive contexts. While this ability is a fundamental component of human musicality it also appears evident in other communicative behaviors and must be considered in the broad context of music as a communicative medium. Functions of human musicality are briefly described and the roles of phase and period correction in human temporal sensitivities and capacities are outlined. The range of temporal behaviors manifested by non-human species is then discussed and compared with those of humans; absence of evidence for period correction in the group temporal behaviors of non-human species suggests that the human capacity to entrain is likely to be species-specific, although research on the rhythmic capacities of non--human species is sparse. A summary of recent experimental research on human temporal interaction indicates that the human capacity to entrain is most acute in contexts that afford mutual temporal adaptation. It is concluded that human capacities for rhythm and entrainment are likely to have emerged initially with the appearance of the *Homo* group proper, around two million years ago, as an exaptive consequence of bipedalism in an increasingly complexly social lineage.

Introduction

From an evolutionary perspective, music poses a puzzle. It is one of the earliest symbolic behaviors for which there exists material evidence in the archaeological record (D'Errico et al, 2003), is found in all known human societies (Blacking, 1995), and members of any given society are generally expected to be able to engage with music in culturally appropriate ways. Yet when compared with other universal human faculties such as that for language, as Jackendoff and Lerdahl (2006) note it seems difficult to conceive of selection pressures that would promote the evolution of a universal human faculty for music, particularly when music is conceived of as complex and pleasing sound patterns that have value only for entertainment.

Nevertheless, when attention is directed to the commonalities across cultures in both the functions that music fulfils and the contexts within which music is typically employed in everyday life, the evolutionary emergence of a faculty for something like music may appear more explicable. Music across cultures is interactive and participatory, is embedded in other activities, and is experienced as having consequences and meanings that extend well beyond activities identifiable as 'musical' (see, e.g., Feld & Fox, 1994; Titon & Slobin, 1996; Nettl, 2005; Cross, 2007). From a cross-cultural perspective music appears more like a communicative medium than a consumable commodity (see Feld, 1996). But as a communicative medium music lacks a crucial feature of the paradigmatic human communicative medium, language; it does not have language's potential to denote and communicate states of affairs unambiguously (see, e.g., Langer, 1942; Meyer, 1956), being unable to express even simple propositions. As Titon and Slobin (1996, p1) note, 'Music is universal, but its meaning is not.' The same piece of music can mean different things in different cultural contexts; it can even mean different things for different participants engaged in the same musical activity; yet

this scope for differences in interpretation does not seem to cause any communicative difficulties for participants.

Indeed, this semantic indeterminacy can be interpreted as enabling music to function as an interactive medium that allows participants to abstract individual and personal meanings without undermining the integrity of a collective musical behavior (Cross, 2001; 2005; 2007). In this respect music appears to have an advantage over language in situations of social uncertainty, and it is in just such situations (ranging in intimacy from mother-infant interaction to national ceremony) where music is most likely to play a significant role in different cultures (see Cross & Woodruff, 2007). Broadly speaking, music can be conceived of as a human communicative medium that has a significant role in the management of social uncertainty.

In fulfilling this role, music does more than provide a semantically indeterminate medium for interaction; it also creates a sense of joint action and common experience by presenting a framework for the periodic coordination of action and perception. Music is generally experienced as organized around a temporally regular series of pulses or beats (which may or may not be represented in the acoustic signal) providing participants in a musical behavior with a common basis for timing their movements and for experiencing musical structures. The human capacity to abstract a regular beat from music and to organize action and perception around the beat seems quite distinct from our ability to process the fine-grained pitch structures that characterize music (Peretz and Hyde, 2003) and has complex neurological correlates (Thaut, 2005). Although it underlies most theories that accord music adaptive value in human evolution on the grounds of its capacity to form or reinforce social bonds (see, e.g., Merker, 2000; Brown, 2000; Dunbar, 2004), its evolutionary roots, and, indeed, its functional scope, are at present obscure. Given the strong likelihood that this capacity plays a significant role in modern human social interaction, it is of considerable importance to arrive at a clearer picture of the human ability to attend to and interact with complexly time-varying patterns of sound and movement.

In humans and other species, the capacity to detect and to integrate events in perception and cognition is governed by temporal limits that have biological bases. Mach, in 1865, observed that 30 msec represents the lowest level for subjective durations, a fundamental time quantum in human perception that has been proposed as deriving from the relaxation oscillations shown by neuronal systems following transduction of a stimulus (Pöppel, 1997). Events require to be separated in time by between 30 and 50 msec in order for discrete events to be detected. Processes of event integration appear equally subject to absolute temporal limits; events separated in time by more than *ca.* two to three seconds are unlikely to be experienced as elements of a single perception-action unit (Fraisse, 1984; Pöppel, 1997). This (roughly) three second limit on event integration in human temporal perception is likely to have origins in common with a generic tendency across a wide range of animal species to produce and to perceive integrated movement patterns with an average duration of around three seconds (Gerstner and Goldberg, 1994; Gerstner and Fazio, 1995)[i]. There thus appear to be continuities between human and non-human temporal capacities, at least in terms of absolute limits on event detection and integration.

There are also, however, significant apparent discontinuities, specifically in respect of the domain of musical rhythm. When presented with a sequence of events which occur at time intervals that are either more or less equal or that conform to multiples or sub-multiples of an equal time interval, a human perceiver will tend to abstract a more or less periodic time interval and organize their attention and action around that interval or *pulse*. This process appears almost automatic for inter-event intervals between *ca.* 200 msec and three seconds (Fraisse, 1982); time intervals consistently beyond these boundaries are simply not accessible to the same processes. Pulses will tend to be experienced as grouped into structures that can be described as meters (London, 2004), with certain pulses within each group appearing more strongly accented than others (Brochard et al., 2003). Significantly, groups of humans can co-ordinate their interactions or *entrain* around a commonly abstracted pulse.

This capacity is not readily observable in the behaviors of other species and may well constitute a species-specific human capacity, a novel ability that arises only in the hominin lineage of the last five million years. If this is so then an understanding of this human capacity to orient attention and action around an abstracted regular pulse and to employ this in social contexts in acts of *entrainment* could be immensely informative in respect of human cultural and communicative capacities in general. However, a caveat must be entered at the outset; large-scale, systematic behavioral observations of pulse-abstraction capacities of other species have not been undertaken. Nevertheless, current knowledge would suggest that reports of periodic behaviors, pulse abstraction or entrainment in other species are motivated by processes quite unlike those that will be described below for humans.

Rhythm and entrainment

Rhythmic behavior appears evident in a wide range of types of interpersonal interaction in which varying degrees of pulse and entrainment create a mutually manifest framework for communication (Clayton et al., 2004; Cross, 2005; Shockley et al, 2003; Richardson et al., 2006). However, entrainment appears most salient in musical activities and a comparative perspective on entrainment in music (Bispham, 2006) highlights some features that appear to identify specifically musical entrainment. These include: the sustained nature of musical pulse; the fact that pulse is perceived unambiguously, or at related hierarchical levels (London, 2004) by enculturated individuals (Stobart and Cross, 2000); and an engagement of the motor system in ways that enable an individual, at least potentially, to manage both fine and gross temporal control in ballistic and smooth movements (Thaut et al., 1997; Bispham, 2003; Thaut, 2005). This last point is strikingly supported by evidence that cerebellarly-damaged subjects who are impaired in the capacity consciously to discriminate rhythms still exhibit the sensorimotor transductive capacities required in synchronous tapping (Molinari et al., 2003).

In human development, it has been claimed that even neonates demonstrate a capacity to entrain with the movements and sounds of a caregiver (Trevarthen, 1999). However, it appears that the ability to sense and respond to pulse of the type evident in music may have a more complex ontogenetic emergence. Bahrick and Lickliter (2004) suggest that infants of around 5 months require multi-modal cues to detect rhythmic changes, whereas 8-month-olds are capable of employing single sensory modalities in such tasks. Phillips-Silver and Trainor (2005) have shown that unless infants of 7 months are moved in time with the meter of a piece of music, they are unable to discriminate between different metric groupings. It appears that a capacity for musical entrainment is consolidated within the first year of life (Hannon and Johnson, 2005; McAuley, Jones et al 2006).

Two distinct correction mechanisms are operational in musical entrainment: phase correction and period correction (Stephan et al, 2002; Repp, 2005). These enable sustained sensorimotor synchronization despite timing errors and motor variance (Wing and Kristofferson, 1973) and musically motivated expressive and structural timing modulations (e.g. Collier and Collier, 2002; Iyer, 2002; Palmer, 1997). Phase correction adjusts for asynchronies between the last response and stimulus events, assuming an unchanged period, whereas period correction modifies the next target interval on the basis of discrepancies between the timekeeper interval and the last or last few inter-stimulus intervals thus altering the period of the attentional musical pulse. In contrast to phase correction, which is most likely generic to all forms of environmental interaction in which short-term future-directed expectancies (following Jones, 1976) are built upon perceived, yet non-isochronous or non-stable, regularities, period correction is observable and functional specifically within the framework of a sustained musical pulse. It may represent a subcomponent of the human ability to set the tempo of a rhythmic activity at will (Repp, 2004). Importantly, period correction has been shown to incur awareness (Repp, 2001) and to be affected by manipulations of attention, intention and awareness (Repp and Keller, 2004) and as such, and again in contrast to phase correction, can be interpreted as a representation of volitional cognitive control.

Rhythm and entrainment in non-human species

Whether or not interactive rhythmic behavior can be considered to be species-specific to humans (apomorphic) or to be shared with related species (synapomorphic) depends largely on the definition used (see Bispham, 2006). Fitch (2006), for example, points to instances of communicatively motivated drumming by palm cockatoos, woodpeckers, kangaroo rats, chimpanzees, bonobos and gorillas. Additionally, distinctive temporal structuring can be used as a source of individuality for sounds not individually marked by the vocal tract (e.g. rhesus macaques screams - Rendall et al., 1998) and animal 'song' in avian (Marler, 2000; Trainer, McDonald & Learn, 2002), cetacean (e.g., humpback whales; Payne, 2000), and pinniped species (e.g., walruses; Sjare et al., 2003; Van Parijs, 2003) clearly involve complex and organized timing. Less convincingly it has recently been claimed that elephants are 'rhythmical' due to their ability to sustain periodic movements with their trunks (Walker, 2001).

It seems most likely, however, that despite some behavioral similarities, none of the above is entirely analogous to musical rhythmic production in humans. Many periodic behaviors (such as occur in elephant trunk swinging) are likely to be simply an outcome of pendular dynamics or biomechanical efficiencies. The complex timing of vocal output (as in animal song) may be entirely the outcome of innate or learnt motor

patterns in which the timing of component events may not be separately controllable, complex sonic temporal patterns resulting from underlying motor programs that are not decomposable by the organism; for instance, the timing of perhaps the most complex example of vocal duetting amongst vertebrates, that of the long-tailed manakin (Trainer, McDonald & Learn, 2002), appears to involve slightly delayed replication of a displaying partner's song rather than any more complex interactive process, albeit that the replication has to be learned. Crucially, no evidence exists in any of the cases cited above that timing or interaction is directed with reference to a future-directed and sustained attentional pulse that characterizes human rhythmical engagement (Jones, 1976).

Comparative research on rhythm perception (Tincoff et al., 2005) has shown that, like human newborns, cotton-top tamarins have the capacity to discriminate between sentences from languages that are differentiable on the basis of their 'rhythm class'[ii]. The hypothesis that differentiation is based on rhythmic cues is supported by evidence that cotton-top tamarins, like newborn babies (e.g. Ramus, 2002) and in contrast to 5-month-old infants (Nazzi et al., 2000), do not differentiate between languages from within rhythmic classes (Tincoff et al., 2005). The possibility that the mechanism underlying our capacity to discriminate languages based on rhythmic cues may have evolved early within the mammalian clade is supported by similar finding with trained rats (Toro et al., 2003). However, as with examples of 'rhythm production' it is likely that results emerge from processes that have little to do with those implicated in musical interaction. The measures that have been employed to characterize different languages according to temporal properties are entirely statistical and capacities to distinguish between 'rhythm classes' are likely to emerge from species-general associative processes rather than from a sensitivity to the fine temporal detail of ongoing event structure.

Whether or not analogies exist, we can at least confidently state that there is something unique about the nature, prevalence, behavioral range and complexity of human rhythmic abilities when compared with those of other species. Existing evolutionary and comparative perspectives on rhythm and pulse in music have pointed to behaviors in the ethological literature (Fitch, 2006; Merker, 2000; McDermott and Hauser, 2005) that are potentially homologous (deriving from a shared lineage) or that are analogous or homoplasies (emerging independently in different lineages in response to similar selection pressures). Attempts have been made to account for the adaptive strength of rhythm and entrainment in the course of human evolution with reference to coalition signaling (Hagen and Bryant, 2003), synchronous mating choruses (Merker, 2000), muscular bonding (McNeill, 1995), mother-infant interaction (Dissanayake, 2000). Hagen and Byrant's hypothesis is that music and dance evolved as a coalition signaling system that could, *inter alia*, credibly communicate coalition quality, thus permitting meaningful cooperative relationships between groups. They argue that this capability may have evolved from coordinated territorial defense signals that are common in many social species, including chimpanzees. While music is seemingly able to fulfill this function it seems unlikely to us that this represents the primary adaptive functionality of rhythm and entrainment. Crucially, it does not account for the full range of entrainment behaviors in humans (ranging from group synchronicity of physicalized action to more subtle manifestations in, for example, mother-infant interactions) and its most prevalent contextualization in affective interpersonal interaction.

Merker (2000) argues that as non-human synchronized group chorusing, exhibited by males of particular Orthoptera (straight-winged insect) and Anuran (frog and toad) species, is displayed exclusively as the outcome of individual mate attraction strategies, it is likely to have been similarly functional during human evolution (though constituting, in the hominin lineage, a homoplasy). However, Merker's argument does not appear to give sufficient weight to the incompleteness of psychological, physiological and behavioral correspondences between the examples given and synchronized behaviors in humans. In contrast to entrainment in humans, correction mechanisms in the animal examples are entirely explicable in terms of phase correction mechanisms to individually specific quasi-eigenfrequencies (Greenfield, 1994) with no evidence for period correction. That is, individuals do not change the baseline tempo of their actions but merely correct for short-term asynchronies with a maintained baseline (cf Vorberg and Schulze, 2002), the temporal structure of signaling is exclusively periodic (i.e. no temporal structuring around the base periodicity), and is, in all cases, restricted to a single modality. Additionally, the fact that these behaviors are manifested solely in the context of male sexual display differentiates them clearly from those involved in human musical entrainment, where there is certainly no evidence for the sexual differences in capacities for entrainment that would be expected were the capacities to be emergent from processes of sexual selection.

More common in non-human species than synchronized group chorusing is duetting, the simultaneous production of coordinated sound sequences by pairs of conspecifics. Duetting appears to offer a paradigmatic case of non-human species communicatively interacting in sound that could inform an

understanding of human rhythmic and entrainment capacities. Many avian duets take the form of antiphonal singing, where the streams of sound events produced by each duetting partner alternate or overlap in time (Mann, Marshall & Slater, 2003). Slater (2000) proposes that the principal function of avian duetting is as a form of cooperative territory guarding (for instances, a female may hold a territory, attracting a male with whom a pair-bond is formed; the male will subsequently contribute a coda to the female's song which can be interpreted as deterring other males from seeking to mate with her, while her song function continues to be territorial). Avian duetting may also serve the function of reinforcing, or even maintaining, a pair-bond, although, as Slater (2000, p58, 'As yet, any possible link between this aspect of birdsong and coordinated singing would be decidedly tenuous!'

More potential for relating duetting capacity to human entrainment might be expected to be found in examples from the primate lineage, where the duetting of gibbons is well-documented (Geissmann, 2000), although primate duetting has a long evolutionary history, appearing in some species of strepsirrhine (Geissmann, 2000; Rasoloharijaona et al., 2006). In the latter, evolutionarily more ancient, group of species, its function is evidently as a cooperative display of territory defense (Rasoloharijaona et al., 2006), while in the former group of species, much more closely related to humans, it has been ascribed a function in both territory defense and pair-bonding (Geissmann, 2000). Duetting in gibbons does not appear to be oriented around a steady pulse, typically showing an increase in rate as a song bout progresses; calls tend to be stereotypical and are not the result of vocal learning, which would suggest that duetting in gibbons is rooted in motor programs that are not decomposable. Whether primate duets occur between sportive lemurs or between gibbons, apart from poorly-documented evidence cited by Geissmann (2000, p119) there is no evidence for any tendency to adhere to regular periodicities, nor is there evidence for entrainment between duetting individuals that involves the mutual adaptation of phase and period of the signals produced. Overall, current theory would suggest that the human capacities for musical, rhythmic, behavior and entrainment may well be species-specific and apomorphic to the hominin clade, though, as noted above, systematic observation of, and experiment on, non-human species' capacities remains to be undertaken.

The nature and functions of human entrainment

At this point it should be noted that almost all evidence for the nature of the processes that subserve the human capacity for musical, rhythmic behavior and entrainment has been obtained within a paradigm of sensorimotor synchronization. In this experimental approach, individual participants tap along an isochronous pacing signal (effectively, a metronome). This research has generated a number of models of timing. Perhaps the most widespread are those which suggest the existence of an internal clock (Wing and Kristofferson 1973; Mates 1994; Aschersleben and Prinz 1995,) with error correction mechanisms that adjust the period and phase of the tapping to keep it in time with the isochronous signal or to respond to perturbations in that signal (Vorberg and Wing 1996; Semjen et al. 1998; see Repp 2005 for review). Another class of model is based on dynamic systems theory and propose that these timing processes are more akin to weakly coupled oscillators (Large et al. 2002; Pressing 1999; Toiviainen and Snyder 2003). While the experiments in the first class deal primarily with finger tapping, the experimental work for the second class has also employed pendulum-swinging.

While there is empirical support for both types of model, neither provides comprehensive foundations for explanations for the human capacity to entrain. Both deal exclusively with situations where one adaptive, responsive agent is reacting to non-adaptive and non-responsive stimuli. However, in interpersonal entrainment, two adaptive and responsive agents appear to entrain with each other. In such situations both agents are adapting to each other's variations in timing even in the absence of pacing stimuli; both can be presumed to abstract a common pulse or beat that serves as virtual pacing signal. The processes that subserve this capacity are complex and have not so far either been empirically explored or satisfactorily modeled; for example, computational attempts to represent inter-agent entrainment such as the phase error correction model of Vorberg and Schultze (2002) have tended to fail by virtue of being unable to adapt to the adaptations of the human counterpart.

There are very few studies where the interaction of two participants has been investigated directly from a cognitive point of view. Mates et al., (1992) investigated tapping accuracy by two participants who occasionally had auditory information of each others' behavior; Boker and Rotondo (2003) examined synchrony and symmetry in dancing; and Schmidt and Turvey (2004) used pendulum-swinging instead of

tapping to explore mutual phase correction in rhythmic interaction. However, recent experiments have directly investigated the importance of mutual adaptability in interpersonal interaction (Himberg and Cross 2004; Himberg 2006; Himberg, forthcoming). In a series of 'co-operative tapping' experiments, two participants engaged in rhythmic production - finger tapping - tasks in real-time interaction. The experiments manipulated whether or not participants could hear a pacing signal through headphones, whether they could see and hear each other, and whether they were required to synchronize with the pacing signal and/or with each other. The results of these experiments have shown that participants strongly prefer to entrain with other humans rather than with a non-responsive pacing signal. There are, of course, socio-psychological reasons for this preference - such as saving face, or reluctance to engage in what their partner might judge to be over-assertive behavior, but it also appears that being entrained to another person is somehow natural and very difficult to resist. Mutual adaptability is evident when participants have access to either or both visual and auditory information about their tapping partner's temporal behavior. However, the provision of both auditory and visual information does not necessarily result in the highest level of performance, which tends to result from access to auditory information only; lowest levels of performance are observed when participants can see but not hear each other (cf Repp and Penel, 2002).

The preference for synchronizing with humans rather than a computer-generated pacing signal can be evaluated by monitoring the synchronization errors (timing differences) between the two tappers and between each tapper and the pacing signal. Even in an experiment where both tappers have been instructed to synchronies with the same, isochronous pacing signal, the participants exhibit smaller synchronization errors in relation to each other than in relation to the pacing stimulus. This occurs even when they are told to focus primarily on the pacing metronome and ignore the other tapper. While most tappers are able to stay in time with both the other tapper and the pacing signal, some pairs 'wander off' the metronome completely while keeping closely synchronized with each other. From a 'clock model' point of view this is surprising, as the isochronous pacing signal is perfectly predictable and should therefore be a strong, stationary target. The other tapper, however, is human and therefore variable in their timing, yet is the preferred target.

As the internal clock models suggest, the variability of the human tapping is not completely random. There is an internal structure to it, and it might even have an intentional component, known as expressive timing, or groove (Madison, 2001; Keil, 1987). When tapping isochronous pulses, participants group the individual beats so that they form metrical, hierarchical structures (Povel and Essens, 1985); in musical performance, these structures serve as guides for the performer in executing deviations from metronomic timing, making the music sound more 'alive' and expressive (Repp, 1998). However, it seems that in these entrainment experiments there may be components of variability that are meaningful but distinct from these involved in expressive timing. In an experiment intended to explore the parameters of adaptability, we substituted from time to time for the human partner an 'averaged' playback of human tapping on computer (Himberg 2006, Himberg, forthcoming). The playback exhibited all the characteristics of human tapping performance including grouping information and expressive timing deviations, but was non-adapting, and we found that participants demonstrated less co-ordination with the recorded tapping than with 'live' human partners. In a follow-up experiment it was observed that participants were able to discriminate between an adaptive human partner and non-adaptive, though mock-expressive, computer.

In these experiments it appeared that participants were highly attuned to the presence of their tapping partners, being sensitive to the degree to which these partners engaged in processes of mutual co-adaptation of phase and period in their tapping to maintain perceived synchrony. These findings raise two questions: why should humans do this, and why should a predisposition towards mutual co-adaptation in the timing of sounds and actions be particularly prominent in music? These questions can be addressed, at least in part, by considering how music appears to constitute a quite specific mode of human interaction, one that is oriented towards the reduction of social uncertainty.

Musical interaction can be conceived of as having as its goal either a particular subjective state or a particular intersubjective end-state. At the level of the individual, music may be engaged with as a means of self-regulating affective states; in a group context, collective musical activities can be viewed as providing a potent means of facilitating the convergence of affective states amongst participants. While humans appear to be intrinsically motivated to share psychological states (Trevarthen and Aitken, 2001) and to have a fundamental need for enduring interpersonal attachments (Baumeister and Leary, 1995), a pre-existing state of 'affective divergence' (interpersonal tension) can be postulated as holding between prospective participants in a collective musical behavior. Interpersonal tension is considered to be biologically disadvantageous in terms of both the immediate and the cumulative consequences of adapting to changes in the social environment. The immediate consequence of social engagement can be thought of as the metabolic load

associated with increases in cardiac output in response to challenges (Bazhenova et al., 2001); cumulative consequences can be referred to the concept of 'allostatic load' (see McEwen, 2001), a measure of the cumulative 'stress' associated with adapting to social and physical environmental change.

Against this background, group engagement in music (particularly in singing) can be thought of as regulating arousal and valence and as promoting positive intersubjective emotion by providing an auditory-motor structure for the interpersonal attentional tracking (Jones and Boltz, 1989) of temporal targets through processes of entrainment. Music regulates affect and promotes positive interpersonal states by enabling a joint focusing of attention on actions (particularly vocal) and cues associated with an intrinsic motivation to achieve intersubjective states, and by generating matches between the actual state of the interactants and the desired, intersubjective end-state. This end-state represents a complete resolution of interpersonal tension (i.e., mismatches between the actual states and the goal state of the interactants), which can be viewed as being coded in terms of varying degrees of mutual deviation from common periodicities in the timing of behavior. The function of group engagement in musical behavior can thus be viewed as that of promoting positive intersubjective emotion, and of minimizing the biological penalties of musically mediated social engagement, a function in which processes of interpersonal entrainment play a crucial role (see also Benzon, 2001).

The evolution of human rhythmic capacities: precursor capacities

In summary, it seems likely that the human ability to perceive and produce rhythm in music is one facet of a broader human capacity to behave periodically in time, which is intimately linked to control of gross and fine movement. This broad capacity for periodic behavior is, in turn, linked to an ability to engage in interactive periodic behaviors with others, using both sound and movement cues to ensure mutual adaptation of temporal behaviors in a wide range of interactive and social contexts. Indeed, humans show a preference for temporal interaction with other humans as opposed to synchronization with a metronomic or time-varying pulse; humans are capable of, and are motivated towards, entraining with each other through processes of mutual co-adaptation of phase and period of action. And musical entrainment can be conceived of as co-ordinating the socio-intentional regulation of interpersonal tension and thus serving to minimize the biological costs of adapting to changes in the social environment. It can be hypothesized that the human capacity for musical rhythm is intrinsically social, and that entrainment provides a scaffolding for joint attention and action.

Given these conclusions, what are the implications for an evolutionary view of musical rhythm? The present state of knowledge suggests that human entrainment processes are different from those exhibited by other species, and it appears that the capacity and motivation to entrain by means of mutual co-adaptation of phase and period may well be novel in the hominin lineage, arising at some time within the last five million years. Nevertheless, it is unlikely that the human capacity to entrain adaptively to a pulse is a monolithically novel trait that appears, fully-formed, only with modern humans. It is more probable that different facets of the modern human facility for entrainment have different and potentially long time-depths in the hominid lineage, with the capacity for fully mutually-adaptive entrainment emerging more recently as a composite or mosaic capacity (Caporael, 2001). Entrainment in modern humans involves mutually-adaptive periodic sound-producing behavior, is grounded in cooperative interaction with others, and is multi-modal, involving regular temporal patterning not only of sound but also action. This suggests that aspects of three distinct behavioral capacities evident in non-human species are implicated in most instances of entrainment in modern human behavior, whether in musical interaction or in more general communicative contexts: a capacity for periodic sound-producing behavior; duetting, the motivation and capacity to interact dyadically with conspecifics in simultaneous or interlocking streams of sound; and the use of gesture in conjunction with vocal signals in communicative interaction. Analogues of some of these capacities are evident in the behaviors of quite diverse species-groups; for instance, plain wrens engage in duets, while both woodpeckers and kangaroo rats appear to drum communicatively. Although these behavioral homoplasies may be informative about the ways in which generic environmental situations may impose selection pressures for specific types of behavioral 'solutions', it appears most relevant to consider possible behavioral homologies.

As Fitch (2006) notes, African great apes engage in periodic drumming behaviors in a variety of contexts. Gorillas will employ bimanual drumming ('chest-beating') in male display, but also in play, and the behavior appears to be engaged in by both females and juveniles in play contexts. Chimpanzees will drum on tree buttresses, or, in captivity, on other sounding surfaces such as hollow walls or up-turned buckets,

typically in male display behaviors. While these behaviors appear to exhibit more-or-less regular periodicities, they have not been observed to occur in group contexts that exhibit any degree of synchrony or of mutual adjustment between participants of period and phase; neither has their internal structure been explored to analyze the extent to which the timing of component events ('beats') may be controlled by the sound producer. Hagen and Hammerstein (2007, in press), in proposing that these display behaviors, involving group vocalizations and drumming, might be homologues of early musical behaviors, having the function of group defense of a territory. However, they note that synchronization is not required in order to fulfill this function; indeed, the extent to which their argument is relevant to an understanding of the evolution of human musical behaviors (and particularly, of entraining behaviors) rather than to the emergence of capacities for generic group display is unclear. On the whole, the periodic sound-producing capacities of great apes do not appear strikingly like the types of entraining behaviors evident in modern humans; nevertheless, these capacities for periodic sonic behavior can be thought of as precursors (both in terms of capacity, and motivation) of aspects of modern human rhythmic behaviors.

As noted above, duetting behaviors in gibbons consist of interaction in sound and action in a communicative context. While gibbon duetting fulfils a function of co-operative territory marking, it also appears to reinforce bonds between duetting pairs, being directed towards maintaining mutually affiliative relationships. Gibbon duets exhibit little sense of the periodicity or timing control that are implicated in human entrainment processes; however, the social functionality of duetting in gibbons appears to fit with the intrinsically social nature of rhythmic entrainment in humans, whether in music or in other communicative contexts, and a generic primate duetting capacity might be considered one possible precursor of modern human entrainment capacities. The multimodal nature of entrainment processes is perhaps reflected in the recent findings that chimps and bonobos - but particularly bonobos - will flexibly combine vocalizations with orofacial and brachiomanual gesture in communicative contexts. Pollick & De Waal (2007, p8188) suggest that 'the bonobos' variable gestural repertoire and high responsiveness to combinatorial signaling may have characterized our early ancestors', and it seems feasible that this capacity for the complex combination of multiple communicative channels can be regarded as a likely precursor of at least some facets of human entrainment capacities.

The evolution of human entrainment: possible scenarios and implications

Two scenarios appear viable in accounting for the emergence of the modern human capacity for entrainment. Dissanayake (2000) has argued that modern human rhythmic abilities are grounded in the provision of a temporal framework for multi-modal mother-infant interactions, in which mothers vocalize to infants using heightened or exaggerated speech (motherese, or Infant-Directed speech) together with rhythmic touching and moving. Empirical research has substantiated the rhythmic and multimodal nature of mother-infant interaction in humans (Papousek, 1996; Trevarthen, 1999; Gratier, 1999; Longhi & Karmiloff-Smith, 2004), and Dissanayake (2000) and Falk (2004) see the emergence of multimodal mother-infant capacities as key to the evolution of, respectively, music, and language. Falk (2004), in a review of non-human-primate and human mother-infant interactive behaviors, notes that our nearest primate relatives do **not** engage in mother-infant interactions of the same types as are evident in all human societies; she proposes that selection pressures accompanying the trend for enlarging brains in late australopithecines/early *Homo* gave rise to the emergence of (p491) 'prelinguistic vocal substrates for protolanguage that had prosodic features similar to contemporary *motherese*', together with increased prevalence of multi-modal communication between mother and infant. In this view, modern human entrainment capacities are likely to have emerged at some point after the *Pan*-human split, though the precise nature of the mechanisms that would have given rise to a prevalence of rhythmic movement in evolutionarily-early mother-infant interactions remains unclear.

An alternative scenario can be proposed which appears to locate the emergence of entrainment capacities more precisely within the hominin lineage, as well as rooting these capacities in broader communicative contexts. The most efficient way of ensuring effective communication is to use all available channels, which, in the case of a social, bipedal hominine, may involve vocal signals, orofacial and brachiomanual gestures (Pollick & de Waal, 2007). If a degree of behavioral similarity between early hominins, such as australopithecines, and extant non-human hominines, such as chimps and bonobos, is assumed, it is likely that australopithecines would have shown a predisposition to employ vocal signals in conjunction with orofacial and brachiomanual gestures. However, for early hominins, bipedal locomotion, most likely arising from pressures on energetic efficiency (Sockol, Raichlen & Pontzer, 2007), could have

limited the number of available communicative channels to the vocal. For gestural channels to be available, each member of a bipedally locomoting group would have to attend visually to each other member, which would have limited the amount of attention available for locomoting and, more crucially, for the detection of threats or of exploitable food resources.

Moreover, communication using vocal sounds and gestures in modern humans appears to require that communicating individuals adopt appropriate positions in respect of each other (the 'F formation', Kendon, 2004), a process that seems incompatible with locomotion. However, modern humans are able to communicate effectively while locomoting. Vocal signals can evidently suffice for effective communication, given the complexity of modern human language, but most ambulating humans also adopt another strategy; they entrain their gait to each other, moving more-or-less synchronously together. This enables the principal visual communicative channels - orofacial, brachiomanual - to be maintained in more-or less consistent locations within each other's visual frame, and can be thought of as a means of creating a 'mobile F formation', effectively maintaining 'readiness to communicate' while co-locomoting.

Readiness to communicate is likely to have been vital for early hominins. As fairly small apes frequenting an open savannah environment that offered little cover, the ability rapidly and effectively - perhaps even silently - to signal danger, or resource availability, to each other would have been at a premium. Were early hominins to have entrained their gait one to another they would have been in a much better position to enact communication effectively (by virtue of ready accessibility of all available communicative channels) than if gait were to have been wholly individual and unsynchronized. Hence it can be hypothesized that there would have been some selection pressure for the adoption of entrainment strategies by bipedal early hominins while locomoting. There would have been immediate costs of engaging in entrained gait, associated with a diminution of bipedal efficiency because of conforming to a gait that might be non-optimal for an individual. However, the fitness of the entraining individual could be considered to have been enhanced by the benefits of a heightened state of communicative alertness, and if this increase in fitness consistently exceeded the costs of entraining, entrainment could become established as an evolutionarily stable strategy.

Although there are, unusually, material traces of bipedal behavior by australopithecines (in the form of the Laetoli footprints: Day & Wickens, 1980), evidence suggests that australopiths were incompletely adapted for bipedalism (Alemseged et al., 2006). In contrast, early *Homo* species (possibly *habilis* or *rudolfensis*, but certainly *ergaster*) were more completely bipedally adapted, showing evidence of selection for not only bipedal walking but also endurance running (Bramble & Lieberman, 2004). It can thus be argued that selection pressures for entrainment may have been most acute for early *Homo*, and it can be hypothesized that with the emergence of early *Homo* a fully-fledged capacity for interactive entrainment enters the human repertoire - with all the costs and benefits that this capacity offered. Outside of a bipedal locomoting context, the costs of initiating an entraining behavior could have been high; the production of periodic movements and sounds that appeared to have a display function (not being oriented either towards defense or aggression) can be though of as having incurred a vulnerability cost (Searcy & Nowicki, 2005). Against this vulnerability would be offset the potentially weighty benefits of the availability of a means of communicating readiness to communicate that was otherwise biologically low-cost. If benefits exceeded costs, entraining behaviors would be likely to become pervasive in communicative behaviors, as appears to be the case for the modern humans. It is notable that behaviors characterizable as synchronization or entrainment appear to emerge as stable strategies for the maintenance of communication in several A-life studies of the evolution of communicative systems (see, e.g., Di Paolo, 2000).

Conclusions

As noted earlier, the human capacity for musical rhythm is intrinsically social, entrainment appearing to provide a scaffolding for joint attention and action. These considerations, together with the likelihood that a fully-fledged capacity for interactive entrainment enters the human repertoire only with the emergence of the genus *Homo*, suggests that entrainment may have played a significant role in the enhancement of perhaps one of the key capacities of the hominin lineage, the capacity for culture (Sperber and Hirschfield, 1999).

Culture has been variously defined in the evolutionary and anthropological literature, but a minimal definition would seem to require the social transmission of information. As van Schalk (2007) notes, an increasing weight of evidence supports the view that the behavior of many species of primate, in particular,

the great apes, exhibits evidence of the social transmission of information. In particular, innovative behaviors have been observed as developing and diffusing within groups of wild chimpanzees in ways that can only be accounted for in terms of social rather than genetic transmission (Whiten et al, 1999). Nevertheless, evidence suggests that the types of inferences typical of human cultural interaction, involving the ascription of attitudes, intentions and goals to others, do not underpin social interaction in chimps. Melis, Hare and Tomasello (2005) showed that chimpanzees are capable of engaging in non-kin collaboration in ways that suggest that they are capable of making inferences about both likelihood and efficacy of cooperation, and suggested (p1300) that 'human forms of collaboration are built on a foundation of evolutionary precursors that are present in chimpanzees and a variety of other primate species'. However, Call & Tomasello (2005) suggest that the extent to which cooperating chimps make inferences about the intentions and goals of others is limited, stating that (p61) '…they [chimpanzees] have the cognitive skills to recall, represent, categorize and reason about the behavior and perception of others, but not about their intentional or mental states…'

Tomasello *et al.* (2005) have proposed that the human capacity for culture is based on a capacity for 'shared intentionality', which incorporates not only a degree of understanding of the intentions of others but also an understanding of the ways in which these relate to common goals in terms of hierarchies of possible actions and potential roles. Such a capacity is present even in young infants yet does not appear to manifest itself in one of our nearest relatives, the chimpanzee (Warneken and Tomasello, 2005). For Tomasello *et al.* (2005, p687), 'The key motivational substrate required for collaboration [of the type that exhibits 'shared intentionality'] is the motivation to share feelings, experiences, and activities with other persons'. However, they do not make any specific suggestions as to what mechanisms might have led to the emergence of positive motivations to 'share intentionality'.

One might view entrainment as a potentially significant contributing factor in this theory of the human capacity for culture. It can be suggested that the motivation to share psychological states could be precipitated and sustained by the guarantee of readiness to communicate implicit in inter-personal entrainment. A proximal function of inter-personal entrainment is the profoundly social one of providing conditions for inferences as to the likelihood of cooperation, lowering the biological costs of interaction and affording a guarantee of co-operativity. As a non-exclusive but foundational sub-component of the human capacities for both musicality and language, entrainment would seem to be an ideal candidate for a mechanism that could have expedited the evolutionary emergence of motivations to act collaboratively in the hominin lineage, facilitating the appearance of the human capacity for culture. Hence the human capacity for entrainment, most salient in the present day in musical rhythm, may well have been an adaptive factor in the evolution of flexible human sociality.

Many of these conclusions must remain speculative on the basis of presently available evidence. More, and more ecologically situated, experimental research is required on the behavioral and neurological substrates of rhythm in human interaction; very few studies to date (other than those instanced above) have explored rhythm as an interactive process, and almost none have explored rhythm in cultural contexts other than that of western societies (amongst the few exceptions are Magill and Pressing, 1997, and Stobart and Cross, 2000). And perhaps even more urgently, rigorous research is required into the capacities of non-human species, particularly primates, for periodic behavior and interactive entrainment. Only then will we be in a position to begin to disentangle the constituents of musicality and identify their relationships with, and their consequences for, other human capacities.

References

Alemseged, Z., Spoor, F., Kimbel, W. H., Bobe, R., Geraads, D., Reed, D., & Wynn, J. G. (2006). A juvenile early hominin skeleton from Dikika, Ethiopia. *Nature, 443*, 296-301.

Aschersleben, G., & Prinz, W. (1995). Synchronizing actions with events: the role of sensory information. *Perception & Psychophysics, 57*, 305-317.

Bahrick, L. E., & Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: A developmental test of the intersensory redundancy hypothesis. *Cognitive, Affective, & Behavioral Neuroscience, 4*, 137-147.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497-529.

Bazhenova, O. V., Plonskaia, O., & Porges, S. W. (2001). Vagal reactivity and affective adjustment in infants during interaction challenges. *Child Development, 72*, 1314-1326.

Benzon, W. (2001). *Beethoven's anvil; music, mind and culture*. New York: Basic Books.

Bispham, J. (2003). *An evolutionary perspective on the human skill of interpersonal musical entrainment*. Unpublished MPhil, Cambridge University, Cambridge.

Bispham, J. (2006). Rhythm in Music: What is it? Who has it? and Why? *Music Perception, 24*, 125-134.

Blacking, J. (1995). *Music, Culture and Experience*. London: University of Chicago Press.

Boker, S. M., & Rotondo, J. L. (2003). Symmetry building and symmetry breaking in synchronized movement. In Stamenov, M. I. & Gallese, V. (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 163–171). Philadelphia: Benjamins.

Bramble, D. M., & Lieberman, D. E. (2004). Endurance running and the evolution of *Homo*. *Nature, 432*, 345-352.

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The 'ticktock' of our internal clock: direct brain evidence of subjective accents in isochronous sequences. *Psychological Science, 14*, 362-366.

Brown, S. (2000). Evolutionary models of music: From sexual selection to group selection. In Tonneau, F. & Thompson, N. S. (Eds.), *Perspectives in Ethology 13: Behavior, Evolution and Culture* (pp. 231-281). New York: Plenum Publishers.

Call, J., & Tomasello, M. (2005). What chimpanzees know about seeing revisited: an explanation of the third kind. In Eilan, N., Hoerl, C., McCormack, T. & Roessler, J. (Eds.), *Joint attention: communication and other minds* (pp. 45-64). Oxford: Oxford University Press.

Caporael, L. (2001). Evolutionary psychology: toward a unifying theory and a hybrid science. *Annual Review of Psychology, 52*, 607-628.

Clayton, M., Sager, R., & Will, U. (2004). In time with the music: The concept of entrainment and its significance for ethnomusicology. *ESEM counterpoint, 1*, 1-45.

Collier, G., & Collier, J. (2002). A study of timing in two Louis Armstrong solos. *Music Perception, 19*, 463-483.

Cross, I. (2001). Music, cognition, culture and evolution. *Annals of the New York Academy of Sciences, 930*, 28-42.

Cross, I. (2005). Music and meaning, ambiguity and evolution. In Miell, D., MacDonald, R. & Hargreaves, D. (Eds.), *Musical Communication* (pp. 27-43.). Oxford: O.U.P.

Cross, I. (2007). Music and cognitive evolution. In Barrett, L. & Dunbar, R. I. M. (Eds.), *Handbook of Evolutionary Psychology* (pp. 649-667). Oxford: Oxford University Press.

Cross, I., & Woodruff, G. E. (2007, in press). Music as a communicative medium. In Knight, C. & Henshilwood, C. (Eds.), *The prehistory of language* (Vol. 1). Oxford: Oxford University Press.

D'Errico, F., Henshilwood, C., Lawson, G., Vanhaeren, M., Tillier, A.-M., Soressi, M., Bresson, F., Maureille, B., Nowell, A., Lakarra, J., Backwell , L., & Julien, M. (2003). Archaeological evidence for the emergence of language, symbolism, and music - an alternative multidisciplinary perspective. *Journal of World Prehistory, 17*, 1-70.

Dawe, L. A., Platt, J. R., & Racine, R. J. (1993). Harmonic Accents in Inference of Metrical Structure and Perception of Rhythm Patterns. *Perception & Psychophysics, 54*, 794-807.

Day, M. H., & Wickens, E. H. (1980). Laetoli Pliocene hominid footprints and bipedalism. *Nature, 286*, 385-387.

Di Paolo, E. (2000). Behavioral coordination, structural congruence and entrainment in a simulation of acoustically coupled agents. *Adaptive Behavior, 8*, 27-48.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interactions. In Wallin, N., Merker, B. & Brown, S. (Eds.), *The origins of music* (pp. 389-407). Cambridge, MA: MIT Press.

Dunbar, R. I. M. (2004). *The Human Story*. London: Faber & Faber.

Falk, D. (2004). Prelinguistic evolution in early hominins: whence motherese? *Behavioral and Brain Sciences, 27*, 491-541.

Feld, S. (1996). Pygmy POP. A genealogy of schizophonic mimesis. *Yearbook for Traditional Music*, 1-35.

Feld, S., & Fox, A. A. (1994). Music and language. *Annual Review of Anthropology, 23*, 25-53.

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition, 100*, 173-215.

Fraisse, P. (1982). Rhythm and tempo. In Deutsch, D. (Ed.), *The Psychology of Music* (pp. 149-180). London: Academic Press.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology, 36*, 1-36.

Geissmann, T. (2000). Gibbon songs and human music from an evolutionary perspective. In Wallin, N., Merker, B. & Brown, S. (Eds.), *The origins of music* (pp. 103-124). Cambridge, MA: MIT Press.

Gerstner, G. E., & Fazio, V. A. (1995). Evidence of a universal perceptual unit in mammals. *Ethology, 101*, 89-100.

Gerstner, G. E., & Goldberg, L. J. (1994). Evidence of a time constant associated with movement patterns in six mammalian species. *Ethology and Sociobiology, 15*, 181-205.

Grabe, E., & Low, E. L. (2001). Durational variability in speech and the rhythm class hypothesis. In Gussenhoven, C. & Warner, N. (Eds.), *Laboratory Phonology, 7* (pp. 515-546). Berlin: Mouton de Gruyter.

Gratier, M. (1999). Expressions of belonging: the effect of acculturation on the rhythm and harmony of mother-infant interaction. *Musicae Scientiae, Special Issue*, 93-122.

Greenfield, M. (1994). Cooperation and conflict in the evolution of signal interactions. *Annual Review of Ecological Systems, 25*, 97-126.

Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Human Nature, 14*, 21-51.

Hagen, E. H., & Hammerstein, P. (2007, in press). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Musicae Scientiae*.

Hannon, E., & Johnson, S. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology, 50*, 354-377.

Himberg, T. (2006). Co-operative Tapping and Collective Time-keeping - differences of timing accuracy in duet performance with human or computer partner. In Baroni, M., Addessi, A. R., Caterina, R. & Costa, M. (Eds.), *Proceedings of the 9th International Conference on Music Perception & Cognition* (pp. 377). Bologna, Italy: 2006 ICMPC.

Himberg, T. (forthcoming). *Cognitive Foundations of Interaction in Musical Time*. University of Cambridge, Cambridge, UK.

Himberg, T., & Cross, I. (2004). Interaction in Musical Time. In Lipscomb, S. D., Ashley, R., Gjerdingen, R. O. & Webster, J. P. (Eds.), *ICMPC 8* (pp. 90). Evanston, Illinois: Causal Productions, Adelaide.

Iyer, V. (2002). Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception, 19*, 387-414.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition, 100*, 33-72.

Jones, M., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*, 459-491.

Jones, M. R. (1976). Time, our lost dimension: towards a new theory of perception, attention and memory. *Psychological Review, 83*, 323-355.

Keil, C. (1987). Participatory discrepancies and the power of groove. *Cultural Anthropology, 2*, 275-283.

Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.

Langer, S. (1942). *Philosophy in a new key*. Cambridge, M.A.: Harvard University Press.

Large, E. W., Fink, P., & Kelso, J. A. S. (2002). Tracking simple and complex sequences. *Psychological Research/Psychologische Forschung, 66*, 3-17.

London, J. (2004). *Hearing in time: psychological aspects of musical meter*. Oxford: O.U.P.

Longhi, E., & Karmiloff-Smith, A. (2004). In the beginning was the song: The complex multimodal timing of mother-infant musical interaction. *Behavioral and Brain Sciences, 27*, 516-517.

Madison, G. (2001). Variability in isochronous tapping: higher order dependencies as a function of intertap interval. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 411-422.

Magill, J. M., & Pressing, J. L. (1997). Asymmetric cognitive clock structures in West African rhythms. *Music Perception, 15*, 189-221.

Mann, N., Marshall-Ball, L., & Slater, P. (2003). The complex song duet of the Plain Wren. *Condor, 105*, 672-682.

Marler, P. (2000). Origins of music and speech: insights from animals. In Wallin, N., Merker, B. & Brown, S. (Eds.), *The origins of music.* (pp. 31-48.). Cambridge, MA: MIT Press.

Mates, J. (1994). A model of synchronization of motor acts to a stimulus sequence I. *Biological cybernetics, 70*, 463-473.

Mates, J., Radil, T., & Pöppel, E. (1992). Cooperative tapping: time control under different feedback conditions. *Perception & Psychophysics, 52*, 691-704.

McAuley, J., Jones, M., Johnstone, H., & Miller, N. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychological Genetics, 135*, 348-367.

McDermott, J., & Hauser, M. (2005). The origins of music: innateness, uniqueness, and evolution. *Music Perception, 23*, 29-59.

McEwen, B. S. (2001). From molecules to mind: Stress, individual differences and the social environment. *Annals of the New York Academy of Sciences, 935*, 42-49.

McNeill, W. (1995). *Keeping together in time: Dance and drill in human history*. Cambridge, MA: Harvard University Press.

Melis, A. P., Hare, B., & Tomasello, M. (2005). Chimpanzees recruit the best collaborators. *Science, 311*, 1297-1300.

Merker, B. (2000). Synchronous chorusing and human origins. In Wallin, N., Merker, B. & Brown, S. (Eds.), *The origins of music* (pp. 315-328.). Cambridge, MA: MIT Press.

Meyer, L. B. (1956). *Emotion and meaning in music*. London: University of Chicago Press.

Molinari, M., Leggio, M. L., De Martin, M., Cerasa, A., & Thaut, M. H. (2003). Neurobiology of rhythmic motor entrainment. *Annals of the New York Academy of Sciences: The Neurosciences and Music, 999*, 313-321.

Nazzi, T., Jusczyk, P., & Johnson, E. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*, 1-19.

Nettl, B. (2005). *The study of ethnomusicology: thirty-one issues and concepts* (2nd ed.). Urbana & Chicago: University of Illinois Press.

Palmer, C. (1997). Music performance. *Annual Review of Psychology, 48*, 115-138.

Papousek, H. (1996). Musicality in infancy research: biological and cultural origins of early musicality. In Deliège, I. & Sloboda, J. A. (Eds.), *Musical beginnings* (pp. 37–55). Oxford: OUP.

Payne, K. (2000). The progressively changing songs of humpback whales: a window on the creative process in a wild animal. In Wallin, N., Merker, B. & Brown, S. (Eds.), *The origins of music.* (pp. 135-150). Cambridge, MA: MIT Press.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Science, 7*, 362-367.

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the Beat: Movement Influences Infant Rhythm Perception. *Science, 308*, 5727.

Pollick, A. S., & de Waal, F. B. M. (2007). Ape gestures and language evolution. *Proceeding of the National Academy of Sciences of the United States*, 9.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Science, 1*, 56-61.

Povel, D. J., & Essens, P. (1985). Perception of Temporal Patterns. *Music Perception, 2*, 411-440.

Pressing, J. (1999). The referential dynamics of cognition and action. *Psychological Review, 106*, 714-747.

Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition, 2*, 85-115.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*, 265-292.

Rasoloharijaona, S., Randrianambinina, B., Braune, P., & Zimmermann, E. (2006). Loud calling, spacing, and cohesiveness in a nocturnal primate, the Milne Edwards' sportive lemur (*Lepilemur edwardsi*). *American Journal of Physical Anthropology, 129*, 591-600.

Rendall, D., Owren, M., & Rodman, P. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (Macaca mulatta) vocalizations. *The Journal of the Acoustical Society of America, 103*, 602-614.

Repp, B. H. (1998). Variations on a theme by Chopin: relations between perception and production of timing in music. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 791-811.

Repp, B. H. (2001). Phase Correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 600-621.

Repp, B. H. (2004). Comments on "Rapid motor adaptations to subliminal frequency shifts during syncopated rhythmic sensorimotor synchronization" by M. Thaut & G. Kenyon. *Human Movement Science, 21*, 61-78.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review, 12*, 969-992.

Repp, B. H., & Keller, P. (2004). Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *The Quarterly Journal of Experimental Psychology, 57*, 499-521.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequence. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1085-1099.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2006). The art of conversation is coordination: common ground and the coupling of eye movements during dialogue. *Psychological Science, 18*, 407-413.

Schleidt, M., & Klein, J. (1997). Segmentation in behaviour and what it can tell us about brain function. *Human Nature, 8*, 7-111.

Schmidt, R. C., & Turvey, M. T. (1994). Phase-entrainment dynamics of visually coupled rhythmic movements. *Biological cybernetics, 70*, 369-376.

Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication; reliability and deception in signaling systems*. Oxford: Princetown University Press.

Semjen, A., Vorberg, D., & Schulze, H.-H. (1998). Getting synchronized with the metronome: Comparisons between phase and period correction. *Psychological Research, 61*, 44-55.

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual Interpersonal Postural Constraints Are Involved in Cooperative Conversation. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 326–332.

Sjare, B., Stirling, I., & Spencer, C. (2003). Structural variation in the songs of Atlantic walruses breeding in the Canadian High Arctic. *Aquatic Mammals, 29*, 297-318.

Slater, P. (2000). Birdsong repertoires: their origins and use. In Wallin, N., Merker, B. & Brown, S. (Eds.), *The origins of music* (pp. 49-64). Cambridge, MA: MIT Press.

Sockol, M. D., Raichlen, D. A., & Pontzer, H. (2007). Chimpanzee locomotor energetics and the origin of human bipedalism. *Proceeding of the National Academy of Sciences of the United States, 104*, 12265–12269.

Sperber, D., & Hirschfield, L. (1999). Culture, cognition and evolution. In Wilson, R. A. & Keil, F. C. (Eds.), *MIT encyclopedia of cognitive sciences* (pp. cxi-cxxxii). Cambridge, MA: MIT Press.

Stephan, K. M., Thaut, M. H., Wunderlich, G., Schicks, W., Tian, B., ., Tellmann, L., Schmitz, T., Herzog, H., McIntosh, G. C., Seitz, R. J., & Hömberg, V. (2002). Conscious and subconscious sensorimotor synchronization - prefrontal cortex and the influence of awareness. *NeuroImage, 15*, 345–352.

Stobart, H. F., & Cross, I. (2000). The Andean Anacrusis? rhythmic structure and perception in Easter songs of Northern Potosí, Bolivia. *British Journal of Ethnomusicology, 9*, 63-94.

Thaut, M., McIntosh, G., & Rice, R. (1997). Rhythmic facilitation of gait training in hemiparetic stroke rehabilitation. *Journal of Neurological Sciences, 15*, 207-212.

Thaut, M. H. (2005). Rhythm, human temporality, and brain function. In Miell, D., MacDonald, R. & Hargreaves, D. (Eds.), *Musical Communication* (pp. 171-191). Oxford.

Tincoff, R., Hauser, M. D., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: further tests with a non-human primate. *Developmental Science, 8 (1), 8*, 26–35.

Titon, J. T., & Slobin, M. (1996). The music-culture as a world of music. In Titon, J. T. (Ed.), *Worlds of music: an introduction to the music of the world's peoples*. New York: Schirmer Books.

Toiviainen, P., & Snyder, J. S. (2003). Tapping to Bach: Resonance-Based Modeling of Pulse. *Music Perception, 21*, 43-80.

Tomasello, T., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and Sharing Intentions: The Origins of Cultural Cognition. *Behavioral and Brain Sciences, 28*, 675-691.

Toro, J., Trobalon, J., & Sebastian-Galles, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition, 6*, 131-136.

Trainer, J. M., Mcdonald, D. B., & Learn, W. A. (2002). The development of coordinated singing in cooperatively displaying long-tailed manakins. *Behavioral Ecology, 13*, 65-69.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae, special issue: Rhythm, musical narrative, and the origins of human communcation*, 155-215.

Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*, 3-48.

Van Parijs, S. (2003). Aquatic mating in pinnipeds: A review. *Aquatic Mammals, 29*, 214-226.

van Schalk, C. P. (2007). Culture in primates and other animals. In Barrett, L. & Dunbar, R. I. M. (Eds.), *Handbook of Evolutionary Psychology* (pp. 103-113). Oxford: Oxford University Press.

Vorberg, D., & Schulze, H. H. (2002). Linear phase-correction in synchronization: predictions, parameter estimation, and simulations. *Journal of Mathematical Psychology, 46*, 56-87.

Vorberg, D., & Wing, A. M. (1996). Modeling variability and dependence in timing. In Heuer, H. & Keele, S. (Eds.), *Handbook of perception and action: Motor skills* (pp. 181–262). London: Academic Press.

Walker, G. (2001). Big band. *New Scientist, 172*, 34-35.

Warneken, F., & Tomasello, M. (2005). Altruistic Helping in Human Infants and Young Chimpanzees. *Science, 311*, 1301-1303.

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E. G., Wrangham, R. W., & Boesch, C. (1999). Cultures in chimpanzees. *Nature, 399*, 682-685.

Wing, A., & Kristofferson, A. (1973). The timing of interresponse intervals. *Perception and Psychophysics, 13*, 455-460.

[i]NOTES

Schleidt & Klein (1997) suggest that different species consistently display different durations of perception-action units (baboons tending to adhere to a one second duration, whereas chimps and humans exhibit units of three seconds' duration); however, irrespective of the absolute durations involved, it appears that the behaviours of a wide range of mammalian species that are oriented towards event integration generally occur within consistent durational limits.

[ii] Different languages exhibit phonological properties that have led them to be classed as having different rhythmic structures; so, for example, French has been claimed to be 'syllable timed', with each syllable being accorded an equal temporal duration, while English is said to be 'stress-timed', with stressed syllables lasting for longer than unstressed ones. This distinction has been challenged by Ramus et al (1999), who suggest that the average durational proportions of vocalic and intervocalic intervals within each language provides a means of differentiating between languages that maps onto measurable temporal properties; Grabe and Lowe (2001) make the alternative proposal that the running average of durations of pairs of syllables (the Pairwise Variability Index) provides a more detailed and secure basis for differentiating between the temporal properties of different languages.